PromptCharm: Text-to-Image Generation through Multi-modal Prompting and Refinement

Year: 2024 | Citations: 107 | Authors: Zhijie Wang, Yuheng Huang, Da Song, Lei Ma, Tianyi Zhang

Abstract

The recent advancements in Generative AI have significantly advanced the field of text-to-image generation. The state-of-the-art text-to-image model, Stable Diffusion, is now capable of synthesizing high-quality images with a strong sense of aesthetics. Crafting text prompts that align with the model's interpretation and the user's intent thus becomes crucial. However, prompting remains challenging for novice users due to the complexity of the stable diffusion model and the non-trivial efforts required for iteratively editing and refining the text prompts. To address these challenges, we propose PromptCharm, a mixed-initiative system that facilitates text-to-image creation through multi-modal prompt engineering and refinement. To assist novice users in prompting, PromptCharm first automatically refines and optimizes the user's initial prompt. Furthermore, PromptCharm supports the user in exploring and selecting different image styles within a large database. To assist users in effectively refining their prompts and images, PromptCharm renders model explanations by visualizing the model's attention values. If the user notices any unsatisfactory areas in the generated images, they can further refine the images through model attention adjustment or image inpainting within the rich feedback loop of PromptCharm. To evaluate the effectiveness and usability of PromptCharm, we conducted a controlled user study with 12 participants and an exploratory user study with another 12 participants. These two studies show that participants using PromptCharm were able to create images with higher quality and better aligned with the user's expectations compared with using two variants of PromptCharm that lacked interaction or visualization support.